Virtual Institute of Microbial Stress and Survival
DOE Genomes To Life Project
Progress Report: June, 2003

I. Overview

The objective of this monthly progress report is to provide an update of the technical and administrative actions from the previous month as well as forecast upcoming progress for the VIMSS Genomes To Life Project. I want to remind everyone how important to make sure everyone is communicating. The discussion boards (http://genomics.lbl.gov/~aparkin/discus) provide a forum for people to ask questions about direction of the project, priorities, and technical issues that can be read and answered by the entire group. I know email is often the most efficient means but it does privatize some of the important communications. Also, posting project data and information to BioFiles (https://tayma.lbl.gov/perl/biofiles) is EXTREMELY important. We are in the process of adding user help files to BioFiles – if you have user questions, please contact Keith Keller (tel: 510.495.2766 or email: kkeller@lbl.gov). This is the best metric I can give to the DOE leadership that we are making progress aside from the VIMSS website. Please make us and yourselves visible by donating data and information to the website.

II. Applied Environmental Microbiology Core

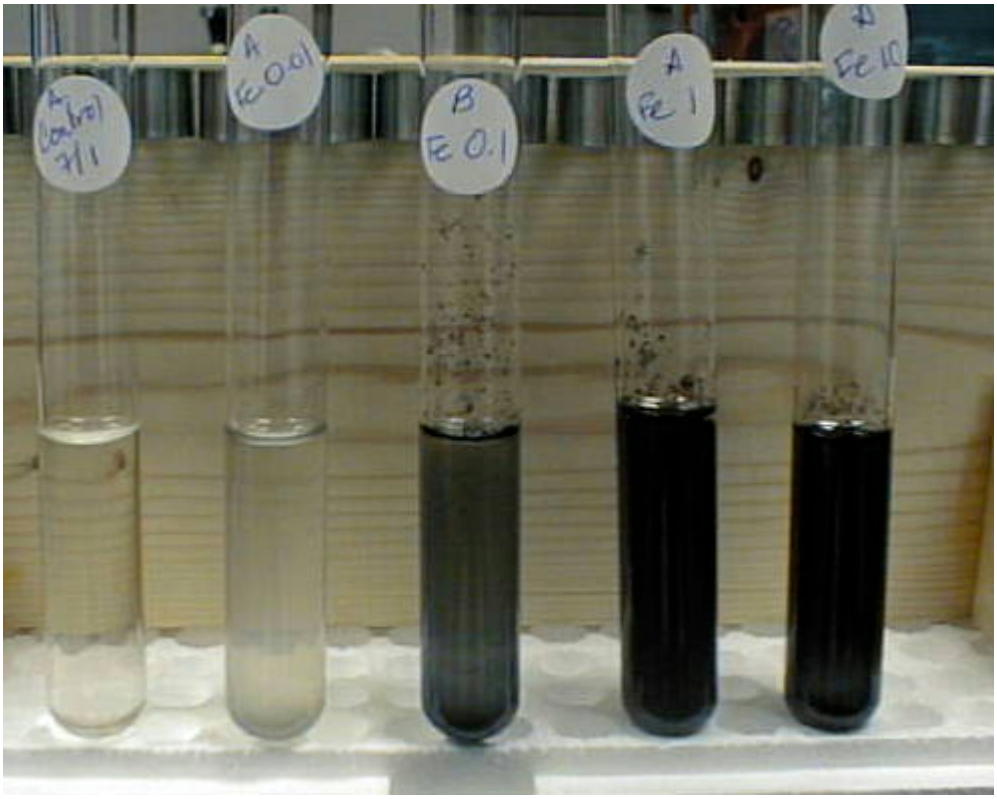
LBNL

SR-FTIR. We continue to modify the existing SR-FTIR spectromicroscopy apparatus to study *Desulfovibrio vulgaris* under anaerobic conditions. Used the anaerobic chamber built in Holman's lab in may to show that Fe, which was used as a bacterial growth indicator in the medium, formed a thick Fe layer on top of the biofilm. These Fe layers interfered with the IR signals. Restarted the study using medium without Fe as a growth indicator.

This month we have collected data on growth curves of *D. vulgaris*. Data is still being processed and will be presented during the poster session at the annual retreat. Due to time constraints during growth curve measurements, rather than doing Live/Dead assays on fresh cells, the cells will be preserved with formaldehyde to perform total cell counts. It appears that stationary phase is reached in Baars medium in about 3 days. Growth in LS medium appears to be slower. We are completing replicate growth curves on different batches medium (both LS and Baars) to insure reproducibility. During growth, the cells are assayed over time using protein assays, OD, total cell counts, pH, and Eh. The growth curves will be repeated with N_2 sparging.

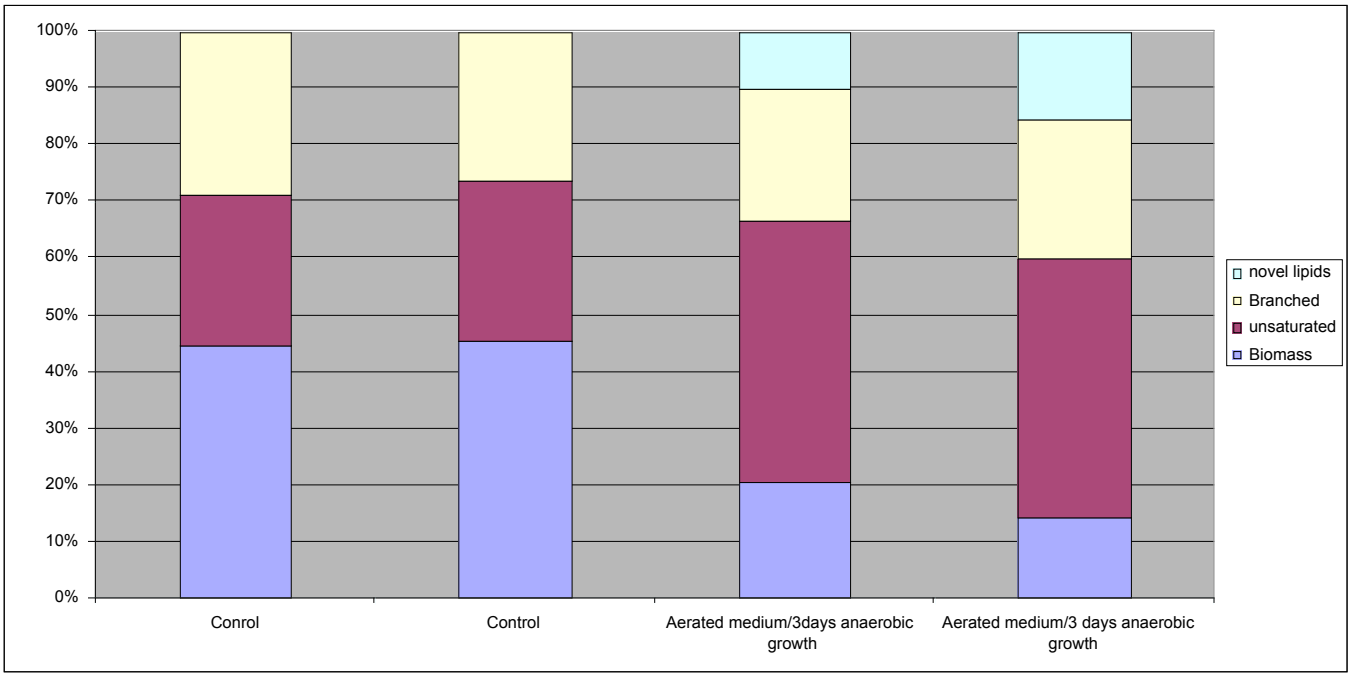
We currently are getting better results with the DC protein assay but will complete the growth curves with both DC and the Bradford protein assay for comparison.

We completed some assays of *D. vulgaris* grown in a range of Fe concentrations, 0 to 10 g/L. Growth is evident at all Fe concentrations. Detailed assessment of the cell density of these cultures is still pending.



Growth of D. Vulgaris in Baars medium in varying Fe concentrations. Left to right: Control, 0.01 g/L, 0.1 g/L,1 g/L, 10 g/L.

Continued lipid analysis of the lipid signatures of oxygen stressed cells included (1) inoculation of aerated No-Fe Baars medium followed by growth in the anaerobic box, and (2) inoculation of No-Fe Baars followed by growth during air sparging. The results of (2) are not complete, and preliminary results from (1) show that new lipids were formed in the oxygenated-medium growth, which could have been either from contamination or from a stress response of the organism.



University of Washington

MPN enrichments of new sediment sample FWB203-03d 04 from FRC area 2 (17.5' to 22.5' depth, sent to us) have been started on media supplemented with one of the following substrates: 1) lactate, 2) acetate, 3) propionate, 4) pyruvate, 5) ethanol and 6) hydrogen with carbon dioxide.

Single colonies have been obtained in agar inoculated earlier with ground water TPB-25Y of FRC area 2 enrichment and demonstrating sulfate reducing activity. Ten were transferred to liquid B3 medium supplemented with lactate and ethanol for further physiological and molecular characterization. However, our other enrichments from sediments FW-109 and 107 of FRC area 3 that showed sulfate reducing activity and were used for inoculation agar roll tubes have not produced visible colonies.

To further characterize metabolic capabilities of model sulfate reducing bacteria we studied growth and metabolite accumulation of two *Desulfovibrio* species (*D. vulgaris* Hildenborough and *Desulfovibrio* sp. PT2)in B3 medium with pyruvate or lactate as carbon and energy sources with no sulfate or other electron acceptor added. Substrate and metabolites were studied using IC Dionex LC 20 equipped with column IonPac AS11. Both strains demonstrated very similar characteristics when fermenting pyruvate. Also some differences in metabolite profiles between these two strains were observed after growth on lactate although compounds corresponding to those peaks have not yet been identified.

As we demonstrated in our previous experiments, 10 g/L of sodium chloride slightly inhibited the growth of both strains in modified McC medium, which we have also used for studies of syntrophic association between these organisms and *Methanococcus maripaludis* (a hydrogentrophic methanogen). We have therefore reduced concentration of NaCl to 2g/l in this medium for continued characterization of syntrophic growth. Currently we are testing this medium for growth *M. maripaludis* and two our model SRB strains.

The FairMenTec Bioreactor has been tested in three batch runs with *Desulfovibrio sp.* PT2 in LS medium supplemented with 20 mM of lactate and 28 mM of sulfate at 30^0C. This strain demonstrated very long lag phase and tendency to attached growth on the walls of the vessel, apparently associated with continuous sparging with nitrogen, When the sparging of nitrogen was discontinued, PT2 grew primarily in suspension and reached an OD 600 of 0.3. The reason for this sparging effect has not yet been determined, but may relate to loss of CO2 or H2 from the medium.

The 16S rDNA and spacer region between 16S and 23S rRNA genes (ISR) have been amplified from 16 Lake Depue isolates and sent to sequencing facility. Sizes of the amplified intergenic spacer region and RFLP analysis demonstrated that at least two different strains have been isolated from Lake Depue metal-contaminated sediments.

Immediate future work

- Continue development of co-culture growth conditions on modified McC medium.
- Obtain and analyze 16S rDNA and dsrAB gene sequences of SRB isolated from Lake Depue metal-contaminated sediments .
- Amplify and sequence 16S rDNA and dsrAB genes from FRC isolates.
- The fixed bed reactor will be tested for its ability to maintain prolonged anaerobic conditions inoculated with population from fresh sediment core from the FRC.
- Continue to test the FairMenTec Bioreactor by running batch cultures of *D. vulgaris Hildenborough* and *Desulfovibrio* strain PT-2.

Oak Ridge National Laboratory

No update.

Diversa

Progress:

- Two suppliers of the DNA polymerase for amplification of gDNA are being compared. The optimal enzyme mix will have low background, will amplify minute quantities of DNA, and will amplify large DNA fragments.

- The ability of the polymerase to amplify "GC-rich" sequences from organisms such as actinomycetes is being confirmed.

- Several libraries have been constructed, and are currently being QC'd for efficiency, insert size, and sequence uniqueness.

Issues:

- Amplification of gDNA before library construction may introduce bias. This will not be important for screening, unless it affects the efficiency of library construction.

Actions:

- Large and small insert DNA libraries are continuing to be constructed for several of the samples from amplified DNA.

- Diversity indexing will be used to evaluate amplification coverage and bias by comparing DNA from direct small fragment extractions, amplified small fragments, and amplified large fragments.

- Work is ongoing to optimize the large insert FACS biopanning protocol in gel microdroplets. Currently, work is being done to optimize the signal-to-noise ratio from experiment to experiment.

III. Functional Genomics Core

Research integration of the core

As the facilities of the Functional Genomics core group are well underway in establishing their respective protocols and methodologies, the first steps at integrating the efforts of the core was initiated. A first experiment was design to investigate stress changes that exposure of *D. vulgaris* to air (O_2–stress). The rationale and objectives of the experiment are described below and details can be found in the protocol posted on the VMISS discussion board

Rationale: Since most contaminated subsurface environments are episodically exposed to oxygenated water via infiltration of snow melt and rain from the surface, and from recharge areas upgradient where O_2 has not been depleted, O_2 is a common stressor for anaerobes like *D. vulgaris* in the environment. In addition, metal and radionuclide reduction by bacteria is dependent upon fairly stringent reducing

(anaerobic) conditions which can be upset by any exposure to oxygen. *D. vulgaris* is tolerant of oxygen via a number of metabolic mechanisms, negative aerotaxis, and via production of EPS. The exact response sequence that *D. vulgaris* and how oxygen affects that response are unknown.

Objectives: The overall objective of this first experiment is to document the array of changes that occur in cells exposed to air for 24h using the array of techniques being developed by the project. The objectives of this experiment are:

1) Define basic O_2 stress changes that occur as measured by the techniques being developed,
2) establish *D. vulgaris* baseline parameters for the techniques being tested,
3) determine logistical constraints for sampling, shipping etc.,
4) establish QA/QC verification of assay protocols,
5) develop functional design criteria for future experiments.

PROGRESS OF EACH FACILITY

Transcriptomics (ORNL)

Objectives

- To perform initial microarray expression profiling studies for the model bacterium *S. oneidensis* MR-1 to establish a baseline response to various environmental stresses.
- To create and use whole-genome microarrays for *Desulfovibrio* and *Geobacter* for analysis of stress-induced transcriptomes.

Progress since last report

- Completed real-time PCR (iCycler) confirmation of a subset of genes showing differential expression based on microarray hybridization: 5 induced genes and 2 repressed genes under pH 4 and 3 induced genes under pH 10. Real-time PCR profiles for all of the selected genes agreed with the microarray results in terms of expression patterns. The writing of a manuscript describing this work also is currently underway.
- Preliminary growth studies looking at the response of wt MR-1 to strontium toxicity have been completed. Preliminary microarray experiments have been completed, and the results are currently being analyzed..
- Completed additional microarray experiments using higher concentrations of hydrogen peroxide for longer treatment periods. Generated an *oxyR*/*dps* double mutant; phenotype characterization of this mutant is currently underway.
- Manuscript describing salt shock response of *S. oneidensis* has been written.

Future work

- Complete pH stress manuscript. Design more detailed time-series experiments.
- Continue microarray expression profiling of MR-1 cells exposed to toxic levels of certain metals (e.g., strontium).
- Target more genes for deletion mutagenesis based on expression data.

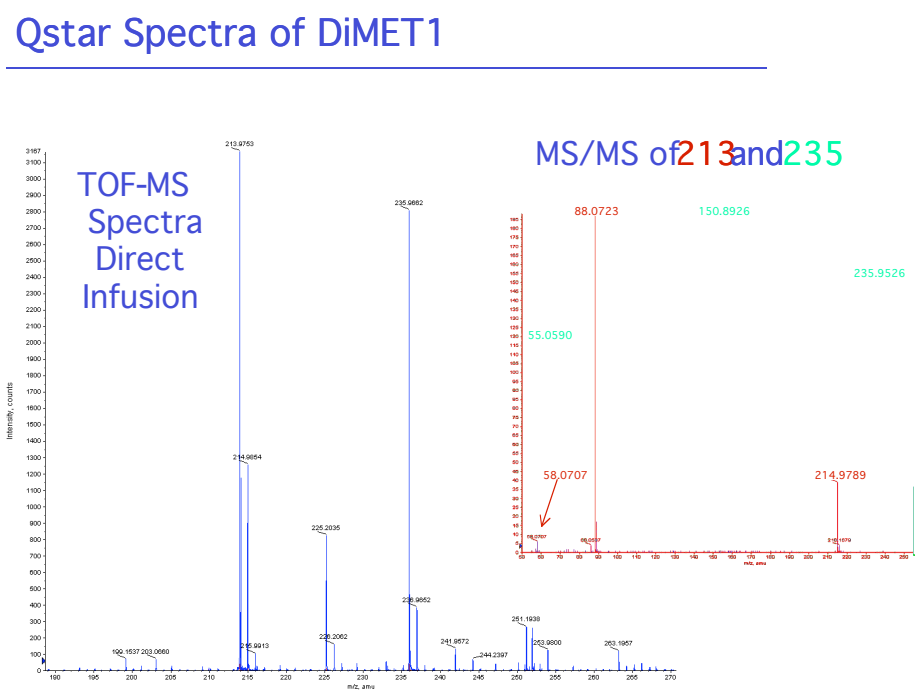
- Construct *Desulfovibrio* whole-genome microarrays.

Proteomics (Diversa)

Overview: Upon the completion of sample preparation optimization and software development, our activities moved towards the method development in quantitative proteomics before more samples arrives.

Objectives
- Data analysis of *D. vulgaris* proteome
- Continue to validate chemical/isotope labeling method(s)

Progress since last report
- So far 1386 total unique proteins identified from *D. vulgaris*
- Method evaluation in DiMET procedures for differential quantitative analysis.
  - Sulfohydral labeling reagents: we have designed several new compounds to label the reduced cysteine side chain to introduce the differential mass. A pair of the compounds DiMET1 and D6 DiMET1) has been synthesized, purified, and analyzed using QSATR (See figure below). We are testing them on standard



peptides, proteins, and cell lysates.

Qstar Spectra of D6 DiMET1

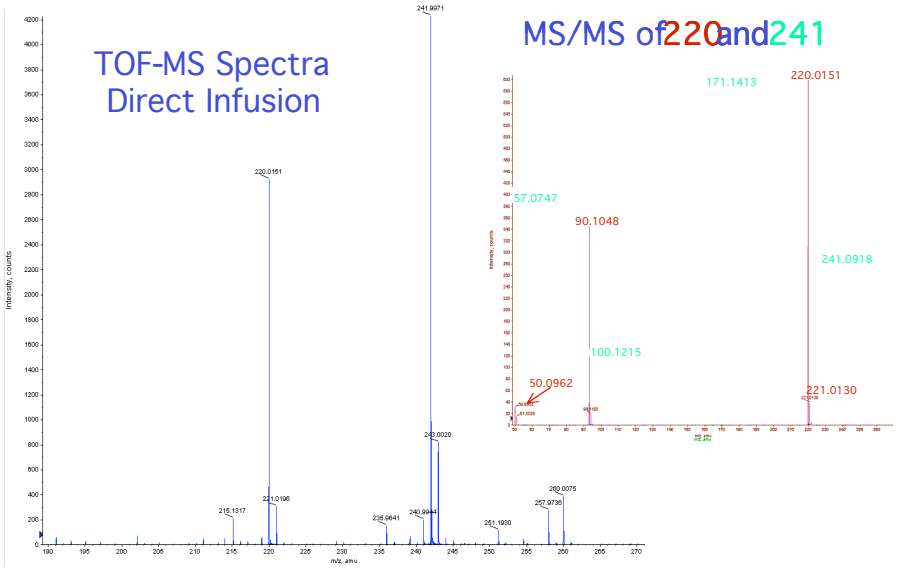


Figure 1. Mass shift analysis produced by DiMET1 labeling cysteine

Protein complexes (Sandia)

Objectives

To work towards isolation of proteins and complexes involved in heat-shock in D. vulgaris and their identification by separation (2D-gel/LC) followed by MS as well as nanoLC/MS/MS of isotope-labeled peptides..

Progress since last report

D. vulgaris Proteomics:

1. Finished analyzing results from ICAT LC/MS runs using improved version of ProICAT at Applied Biosystems. Currently evaluating results

2. Obtained in-gel tryptic digests for the two-dimensional differential in-gel electrophoresis runs on the samples, as used for ICAT, were used in these gels. Currently working on MALDI-TOF runs for these tryptic digests to identify suitable targets. (will be presented at the Annual Retreat and posted on VIMSS web-site).

3. Obtained Speed Vac system from Thermo Savant dedicated for ICAT experiments.

Mass Spectrometry for Protein Complex Identification:

1. Have isolated complexes associated with HSP 70 in D. vulgaris using antibodies attached on magnetic beads

2. Have set up a nanoLC system, including nanoESI-MS interface. Reverse phase

columns are now successfully being packed "on site" for nanoLC/MS as we didn't get satisfactory results with commercially-available columns

3. Demo'd 3 MALDI-TOF instruments. The vendor and price for MALDI-TOF instrumentation has been set. Currently working on buying one.

4. Set up and successfully tested (with digests of standard proteins) HPLC for strong cation exchange chromatography of complex peptide mixtures. Will be used in conjunction with ICAT experiments to improve sample coverage.

Future work

1. Finish nano-LC-MS (QTOF) set-up and optimize ICAT experiments for whole cell protein extracts.

2. Setup MALDI-TOF for 2D gel electrophoresis.

3. Optimize immunoprecipitation experiments for protein complex identification.

Metabolomics and Proteomics (UCB, LBNL)

Objectives

- Optimization of the hydrophilic interaction chromatographic methods for nucleotide separation
- Development of a protocol to extract cell lysis of *D. vulgaris*
- Apply the separation methods for CoAs and nucleotides to *D. vulgaris* cell lysates.
- To create a growth curve for *Desulfovibrio*, incorporating both OD and protein assay at 30°C.
- Test the knockout strategy that was designed for the generation of knockout mutants.
- Complete the first collaborative ICAT O_2 stress response test.

Progress since last report

Metabolite analysis

- The conditions for nucleotide separation were obtained. An Amide-80 column (TOSOH Biosep, 2mm □ 250 mm) is used with a LC Packings Capillary LC system and a QTrap tandem mass spectrometer as the detector. The LC conditions are: 10 minutes of an isocratic elution of 20% buffer (NH_4OAC, 5 mM, pH 5.0) followed a gradient elution of 20-40% buffer in 20 minutes, then another isocratic elution of 40% buffer for 20 minutes.
- Due to the formation of FeS during *D. vulgaris* cell growth, use of silicon oil to separate cells from the medium does not work well. Therefore, cells were harvested, wash quickly, then spun down again followed by suspending them in 10% TCA to obtain cell metabolites. Further optimization is under progress.
- The cell extract obtained by the above method was subject to CoA and nucleotide analysis. CoA, acetyl-CoA, succinyl-CoA, Propionyl-CoA and malonyl-CoA can be detected from the cell extract (Figure 2). However, the analysis of nucleotides is not

very successfully since only a few nucleotides can be detected and AMP is significant compared to other nucleotides.

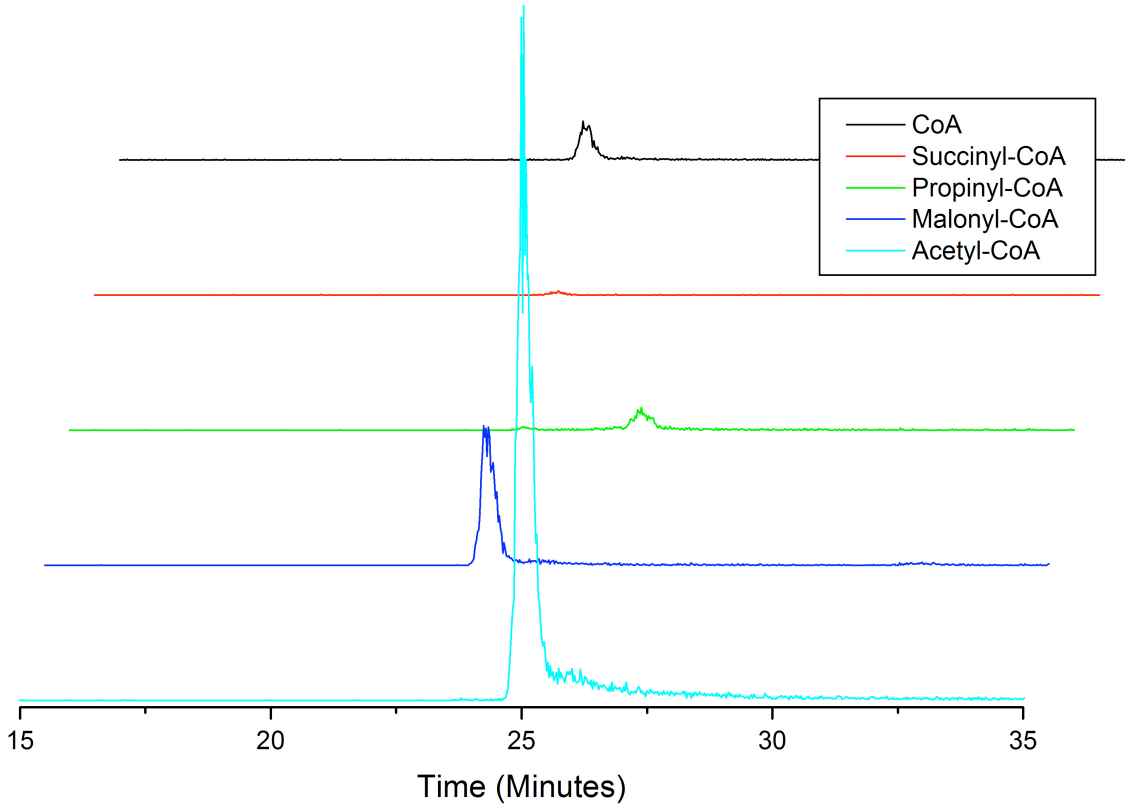


Figure 2. Analysis of CoAs from *D. vulgaris* cell lysate.

Gene knockouts and analysis of two-component systems

- An OD and protein assay growth curve was completed for *D. vulgaris*.
- Conditions for creating the PCR products of interest for the knockout mutants are still being optimized, but progress has been made with some of the open reading frames.
- Towards the goal of creating a library of knock-out mutants for all the two component systems in *D. vulgaris* a few candidates have been chosen to optimize the high throughput strategy outlined in the previous report.
- As reported earlier, HSKs, TIGR ORF00061 and TIGR ORF05707 have been amplified from the genome and cloned into the expression vector pET30a(+). Now, PCR primers to create the open vectors containing these genes were designed and obtained. The primers are designed, to not only “open” the vector, but also to delete a portion of the gene, introduce UP and DOWN TAGs, and create overhangs.

- PCR primers to create antibiotic resistance cassettes with complimentary overhangs that can be annealed into the open vectors were also designed and obtained. PCR conditions for the above mentioned vectors and inserts have been optimized. Procedures to create the necessary suicide vector and its transport into *D. vulgaris* are in progress.

- Furthermore, the required PCR primers for all Histidine Sensor Kinases in the TIGR annotated FASTA files have been designed.

Proteomics

ProICAT software:

The previously unresolved software glitch in ProICAT which resulted in incorrect ID of Heavy tagged peptides has been resolved with the help of the vendor, Applied Biosystems. Existing ProICAT data for sporulation related events at t=30mins in *B. subtilis* have been reevaluated and will be posted on the VIMSS site.

ICAT exprements in *B. subtilis*:

A time course ICAT experiment has been done for sporulation initiation in *B. subtilis* wherein the control proteome (t = -1, before suspension in sporulation media) was tagged and compared with tagged proteomes at time points t=30mins, t=1.5hours and t=3hours after resuspension in sporulation media. MSMS analyses of these samples are in progress.

Future work

- Further optimization of the cell lysis protocol.
- Nucleotide analysis of cell extract
- Development of methods for redox cofactors separation
- Amino acids analysis
- Finish the knockout mutant strategy by completing the annealing and transformation of the vector into *E. coli*.
- As primers arrive, clone the rest of the histidine sensor kinases out of the genome. Work out conditions for transferring plasmids into *D. vulgaris* in our lab.
- Complete a preliminary ICAT oxygen stress experiment with a sample grown in our own laboratory, to validate our method before t

IV. Computational Core

LBNL – Arkin

In June, we completed an update to our comparative genomics database that includes 126 complete and partially sequenced genomes. We completed all-against-all BLAST searches between all protein coding regions in this larger database, and have begun to rebuild our bioinformatics tools to take advantage of this recent influx of data. We have also begun work on an automated pipeline for data input in anticipation of receiving an additional 600 genomes over the next year. Finally, we are setting up a quality assurance pipeline to verify the accuracy of the large amount of sequence and annotation data contained within the database.

We have started development of tools to compare and integrate data from the Functional Genomics Core with our own preliminary annotations. In particular, we are focused on developing a Perl codebase for microarray analysis. We have achieved a flexible schema for inserting raw and/or processed microarray data into our comparative genomics database, and will begin data entry from our literature-culled database shortly. Finally, we are starting to analyze these data by comparison with our predictions of operon, regulon, and cis-regulatory structure, in order to test both the quality of our predictions and to learn which experimental designs yield the greatest quantity of information about gene regulation.

In preparation for the August 15th release of the VIMSS Comparative Genomics Web Tools, we have made progress on our internal development version of the website. We have developed a protein annotation browser that displays all of our internal annotations, a navigable view of the operon structure predicted by our operon calling tool, and a graphical view of the domain structure within each ORF including links to external sites (COG, Pfam, SMART) with more information on each domain. We will integrate these protein pages into our comparative genome browser during the next month.

LBNL – Olken

Biopathways Graph Data Manager

Data Modeling: Contexts

We met with Terry Hazen (LBNL) to discuss our efforts in data modeling of contexts. Hazen thought the contexts would be useful, but was uncertain how they would be queried. Hazen noted that in reality we only know individual context points rather than regions. Points correspond to individual experiments – regions are inferences under the assumption that environmental domains of reactions (gene expression) are continuous and convex. Hazen also noted that we would need to be able to specify extracellular contexts, i.e., geochemical contexts. Hazen was particularly interested in the specification of oxygenation contexts, particularly temporal aspects of stress response to the presence of oxygen. We have not yet included temporal specifications into to our modeling of contexts.

It is increasingly clear that separation of contexts (i.e., their reification) will permit use of a different data model and for contexts (cf. to graph data model for pathways) and possibly a different query processor (e.g., perhaps a logic DBMS or constraint DBMS). Thus efforts to descibe environmental contexts as convex polytopes would probably entail the use of a constraint DBMS for context query processing. Distinct contexts will also permit more concise description of contexts and more efficient processing of context queries (assuming that many reactions use the same contexts), since separate (reusable) contexts should require fewer context comparisons (with possibly some additional list processing).

Implementation Planning
Kevin continued work on the specification of the Navigational API for access to the graph database from Java applications codes. Kevin did a test of the proposed Navigational API. This entailed extraction of the BioDB relational schema via JDBC and schema vixualization with ATT's GraphViz program.

Personnel
We posted the second programmer position for the project to LBNL job listings.

Conferences
We attended SIGMOD Conference and Web DB workshop. Kevin attended the Data Mining workshop. We met with Peter Mork (Univ. of Washington, Seattle) and discussed possible collaboration on RDF-based data exchange and data modeling for biology. We also met with Vassilis Christophides (Univ. of Crete) and discussed possible use of his RDFSuite of RDF query tools.

Publications
Olken (a stand-in for Rotem) presented a joint paper by Frank Olken and Doron Rotem, entitled "Workflow Execution History Data Management: A Framework" at the International Conference on Web Services. This paper discusses a graph data model for workflow execution history data, and transformations between the workflow process model graph and the history graph. In a biological setting the workflow process would be a molecular biological lab protocol (e.g., an automated sequencing protocol, etc.) and the history would be a log of laboratory activities. The paper is relevant to efforts to construct formal models for the representation, storage, vizualization, and querying of laboratory protocols. It could also be useful in managing computational workflows of large numbers of biological datasets (or computations). [Conference attendance was financed by DOE support of Scientific Data Management Group not VIMSS GTL.]

Future work
1) Complete draft of navigational API for the graph DBMS.

2) Elaborate design of context annotations and queries for biopathways (sub)graphs.

3) Commence design of visual graph-based query language.

4) Complete workshop report from NLM Workshop on Data Management for Life Sciences Research.

5) Commence advertising and interviewing for programmer position.

6) Complete P2R for Kevin.

7) Complete Biopathways DB section of DOE Computational Biology Primer for Ying Xu.

8) Work on slides for Graph DB tutorial at IEEE CS Bioinformatics Meeting.

V. Project Management

Project Schedule

The GTL project schedule is undergoing updates and will be posted to the VIMSS Discussion Board on a regular basis. Any updates/comments/revisions should be sent to Nancy Slater via email (naslater@lbl.gov).

GTL Project Meetings

- The next GTL Executive Committee Meeting will be held on September 3, 2003.

- The 2003 VIMSS/GTL Annual Retreat will be held on August 1-3 at the Hotel Nikko in San Francisco. Friday's agenda includes a poster presentation by each of the GTL project team members. The PIs will present on Saturday. See the Retreat website for the registration form and more details (http://vimss.lbl.gov/gtl_ann_retr.html).